ANNALES
UNIVERSITATIS MARIAE CURIE-SKŁODOWSKA
LUBLIN—POLONIA
VOL. XXIX/XXX, 23 SECTIO H 1995/1996

Zakład Nauk Ekonomicznych
Filii UMCS w Rzeszowie

Elżbieta MAKSYMIAK

Minimalizacja efektu katalizy w metodach doboru zmiennych

Minimalization of the Catalysis Effect in the Methods of the Choice of Variables

Zagadnienie doboru zmiennych w jednorównaniowych liniowych modelach ekonometrycznych było przedmiotem uwagi licznych autorów. Stąd w piśmiennictwie naukowym istnieje wiele artykułów, jak też opracowań książkowych, z których najważniejsze to [4], [5], [11]. Do metod doboru zmiennych najczęściej stosowanych należą między innymi: metoda selekcji *a posteriori*, metoda selekcji *a priori*, metoda regresji krokowej, metoda Allena, metoda Hellwiga (por. [4], metoda suboptymalnej wartości współczynnika integralnej pojemności informacyjnej (por. [9]), metoda Rogowskiego (por. [10]), metoda Guzika ([6]). Dobierając zmienne do liniowego jednorównaniowego modelu ekonometrycznego powinno mieć się na uwadze takie postulaty jak:

1) maksymalizację stopnia dokładności, z jaką model opisuje dane zjawisko,
2) istotność parametrów strukturalnych modelu,
3) dokładność ocen parametrów strukturalnych modelu,
4) koincydencję modelu,
5) brak zmiennych współliniowych,
6) brak efektu katalizy lub minimalne jego natężenie.

Metoda selekcji *a posteriori*, metoda selekcji *a priori*, metoda regresji krokowej, metoda Allena, metoda Hellwiga oraz metoda Guzika nie zapewniają koincydencji modelu. W pracy [2] przeprowadzono adaptację czterech pierwszych metod w celu uzyskania przy ich pomocy modeli koincydentnych. Adaptacja tych metod polegała na wkomponowaniu do alogrytmów właściwych tym metodom warunku zapewniającym koincydentność modelu. Uzyskano to poprzez zastosowanie współczynników korelacji częstkowej do badania koincydentności, gdyż znak oszacowania pa-

rametru strukturalnego modelu jest identyczny ze znakiem współczynnika korelacji cząstkowej.

Model otrzymany metodą Rogowskiego oraz metodą suboptymalnej wartości współczynnika integralnej pojemności informacyjnej jest koincydentalny. Żadna z wyżej wymienionych metod nie uwzględnia problemu efektu katalizy.

W niniejszej pracy przedstawimy dwie metody doboru zmiennych, w wyniku których otrzymany model jest koincydentny o minimalnym natężeniu efektu katalizy. Efekt katalizy został wykryty i opisany przez Z. Hellwiga w pracy [7]. Poniżej przypomnimy definicję efektu katalizy.

DEFINICJA 1 (por. [7])

Mówimy, że w modelu ekonometrycznym określonym przez regularną parę korelacyjną (R, R_o) występuje efekt katalizy, jeżeli istnieje taka para wskaźników (i, j), dla której

$$r_{ij} < 0 \quad \text{lub} \quad r_{ij} > \frac{r_i}{r_j}$$

gdzie $R = [r_{ij}]_{k \times k}$, $R_o = [r_i]_{k \times 1}$ są odpowiednio macierzą korelacji dla zmiennych objaśniających i wektorem, którego i-tą współrzędną jest współczynnik korelacji między i-tą zmienną objaśniającą a zmienną objaśnianą. Ponieważ efekt katalizy może występować w modelu z różnym natężeniem, dlatego został określony miernik (por. [7]).

$$\eta = r^2 - H,$$

gdzie r^2 jest kwadratem współczynnika korelacji wielorakiej, zaś H jest pojemnością integralną informacji służący do mierzenia natężenia efektu katalizy. Na podstawie nierówności

$$0 \leqslant \eta \leqslant 1$$

wykazanej w [8] widzimy, że w modelu nie występuje efekt katalizy gdy $\eta = 0$, natomiast natężenie efektu katalizy osiąga wartość największą, gdy $\eta = 1$. W metodach, które zaproponujemy w niniejszej pracy wykorzystamy następującą definicję 2 oraz twierdzenie 1.

DEFINICJA 2 (por. [3]).

Funkcję f określoną na parze korelacyjnej (R, R_o) i przyjmującą wartość w przedziale $\langle 0,1 \rangle$, która jest ciągła względem wszystkich elementów macierzy R oraz wszystkich współrzędnych wektora R_o oraz taką, że gdy R jest macierzą jednostkową stopnia k oraz $\sum_{i=1}^{k} r_i^2 = 1$, to $f(R, R_o) = 1$ nazywamy miernikiem liniowego jednorównaniowego modelu ekonometrycznego.

TWIERDZENIE 1 (por. [3])

Jeżeli f_1 oraz f_2 są miernikami modelu w snsie definicji 1, to dla dowolnych liczb rzeczywistych dodatnich c_1 oraz c_2 funkcja $f = f_1^{c_1} f_2^{c_2}$ jest miernikiem modelu.

Zauważmy, że w sensie definicji 1 miernikami modelu są:

— integralny współczynnik pojemności informacji H,

— integralny współczynnik koincydencji $M^2 = r^2 \cdot \det R$ zdefiniowany po raz peirwszy w [1],

— współczynnik $K = 1 - \eta = \varphi^2 + H$ ($\varphi^2 = 1 - r^2$ jest współczynniem indeterminacji) mierzący natężenie efektu katalizy w ten sposób, że gdy w modelu brak jest efektu katalizy to $K=1$, natomiast wraz ze wzrostem natężenia katalizy współczynnik K maleje od 1 do 0; wartość 0 przyjmuje dla największego natężenia efektu katalizy.

W myśl twierdzenia 1 z mierników H, M^2 i K tworzymy następujące dwa mierniki:

$$H_K = H \cdot K,$$
$$N = M^2 \cdot K$$

które są miernikami modelu. Wiadomo, że miernik H przy doborze zmiennych do modelu uwzględnia zarówno korelację między zmienną objaśnianą a zmiennymi objaśniającymi, jak też wzajemną korelację między zmiennymi objaśniającymi. Współczynnik $H=1$ wtedy i tylko wtedy, gdy $r^2 = 1$ oraz $R = I$, gdzie I jest macierzą jednostkową. Wartość $H=0$ oznacza brak jakiejkolwiek korelacji między zmienną objaśnianą a zmiennymi objaśniającymi. Ponieważ w modelu powinny występować takie zmienne objaśniające, dla których wartości H, K i M^2 są możliwie duże dlatego pożądana jest również duża wartość miernika H_K oraz N.

Poniżej przedstawimy pierwszą z zapowiedzianych metod. Niech $\chi = \{X_1, X_2, \ldots, X_K\}$ będzie zbiorem potencjalnych zmiennych objaśniających. Przez χ_i $(i = 1, 2, \ldots, 2^K - 1)$ oznaczmy i-ty podzbiór zbioru χ. Dla każdego podzbioru χ_i $(i = 1, 2, \ldots, 2^K - 1)$ obliczamy współczynnik H_K oznaczając go symbolem $H_K(\chi_i)$. Następnie postępujemy według algorytmu:

1) Znajdujemy zbiór χ_{i_1} spełniający warunek

$$H_k(\chi_{i_1}) = \max_{1 \leqslant i \leqslant 2^k - 1} \{H_k(\chi_i) : \chi_i \subset \chi\}$$

2) Sprawazamy, czy model o zmiennych objaśniających należących do zbioru χ_{i_1} jest koincydentny. Jeśli tak to procedura jest zakończona i do modelu zostają wybrane zmienne objaśniające ze zbioru χ_{i_1}. Jeśli nie to

3) Wyznaczamy zbiór χ_{i_2} taki, że

$$H_k(\chi_{i_2}) = \max_{1 \leqslant i \leqslant 2^k-2} \{H_k(\chi_i) : \chi_i \neq \chi_{i_1} \text{ i } \chi_i \subset \chi\}$$

4) Sprawdzamy, czy model, w którym zmienne objaśniające należą do zbioru χ_{i_2} jest koincydentny. Jeśli tak, to kończymy algorytm i za zmienne objaśniające do modelu przyjmujemy zmienne ze zbioru χ_{i_2} Jeśli nie, to

5) Wyznaczamy zbiór χ_{i_3} spełniający warunek

$$H_k(\chi_{i_3}) = \max_{1 \leqslant i \leqslant 2^k-3} \{H_k(\chi_i) : \chi_i \neq \chi_{i_1} \text{ i } \chi_i \neq \chi_{i_2} \text{ i } \chi_i \subset \chi\}$$

6) Postępujemy analogicznie jak w 4) itd.

Koincydentność modelu sprawdzamy wykorzystując następujący warunek konieczny i wystarczający na to, by dowolna zmienna objaśniająca miała własność koincydencji.

TWIERDZENIE 2 (POR. [9])

Zmienna objaśniająca X_i jest koincydentna w modelu określonym przez regularną parę korelacyjną (R, R_o) wtedy i tylko wtedy, gdy

$$r_1 > [R_i^i]^T [R_{ii}]^{-1} R_o^i$$

gdzie R_i^i oznacza i-tą kolumnę macierzy R z pominięciem i-tej współrzędnej, R_o^i powstaje z wektora R_o przez odrzucenie i-tej współrzędnej, natomiast R_{ii} oznacza podmacierz otrzymaną z macierzy R przez skreślenie wiersza oraz kolumny o numerze i-tym.

Model jest koincydentny, jeżeli każda zmienna objaśniająca tego modelu ma własność koincydencji. W celu łatwiejszego korzystania z twierdzenia 2 proponujemy wykorzystać następujące.

TWIERDZENIE 3

Jeżeli na elementach macierzy A_1 postaci

$$A_1 = \begin{bmatrix} A & f \\ g^T & z \end{bmatrix}$$

gdzie $A_1 = [a_{ij}]$, $f = [f_i]$, $g = [g_i]$ $(i, j = 1, 2, \ldots, k)$, $z \in R$ (R jest zbiorem liczb rzeczywistych) wykonamy takie przekształcenia elementarne (rozpoczynając od elementu a_{11}), które sprowadzają ją do postaci górnej macierzy trójkątnej $D = [d_{ij}]$, to zachodzi równość

$$Z - g^T A^{-1} f = d_{k+1, k+1}.$$

dla macierzy A_i postaci

$$A_i = \begin{bmatrix} R_{ii} & R_o^i \\ [R_i^i]^T & r_i \end{bmatrix}$$

Na elementach macierzy A_i zgodnie z twierdzeniem 3 wykonujemy odpowiednie przekształcenia elementarne spowadzające ją do postaci macierzy trójkątnej $D_i = \left[d_{pg}^{(i)}\right]$ (p, g=1,2, ..., k). Jeżeli $d_{kk}^{(i)} > 0$ to zmienna X_i jest koincydentna w modelu określonym przez parę korelacyjną (R, R_o). Jeżeli natomiast $d_{kk}^{(i)} < 0$ to zmienna X_i nie jest koincydentna. Model, który otrzymamy w wyniku opisanej wyżej procedury ma największą wartość miernjka H_k wśród modeli koincydentnych o zmiennych objaśniających należących do podzbiorów χ_i (i=1, 2, ..., 2^k-1).

Druga proponowana metoda wykorzystuje integralny miernik koincydencji M^2 oraz następujące

TWIERDZENIE 4 (POR. [1])

Zmienna objaśniająca X_i (i=1, 2, ... „ k) w modelu określonym przez regularną parę korelacyjną (R, R_o) jest koincydentna, jeśli spełniona jest nierówność

$$M^2 > M_j^2$$

gdzie M_j^2 jest integralnym miernikiem koincydencji modelu, który powstaje z wyjściowego modelu przez odrzucenie j-tej zmiennej objaśniającej. W metodzie tej najpierw wyznaczamy wszystkie podzbiory zbioru $\chi = \{X_1, X_2, \ldots, E_k\}$ (jest ich 2^k-1) i dla każdego z nich obliczamy wartość miernika N. Wartość miernika N dla i-tego podzbioru χ_i oznaczamy symbolem N (χ_i).

Procedura postępowania jest następująca:

1) spośród wartości N (χ_i) (i=1, 2 ..., 2^h-1) wybieramy wartość największą. Podzbiór zbioru χ, któremu odpowiada ta największa wartość oznaczamy symbolem χ_i czyli mamy nierówność

$$N(\chi_{i_1}) \geqslant N(\chi_i),\ i = 1, 2, \ldots, 2^k - 1$$

2) Sprawdzamy, czy zachodzi nierówność

$$\bigwedge_j M^2(\chi_{i_1}) > M_j^2(\chi_{i_1}), \text{gdzie } M^2(\chi_{i_1})$$

jest integralnym miernikiem koincydencji dla modelu o zmiennych objaśniejących należących do zbioru χ_{i_1} natomiast $M_j^2(\chi_{i_1})$ jest integralnym miernikiem koincydencji obliczonym dla modelu o zmiennych objaśniających ze zbioru χ_{i_1} bez j-tej zmiennej objaśniającej. Jeżeli tak, to postępowanie kończymy i za optymalny zbiór

zmiennych objaśniających uznajemy zbiór χ_{i_1} Otrzymany model ma własność koincydencji w myśl twierdzenia 4. Jeżeli nie, to

3) Spośród wartości N (χ_i) dla $i=1, 2, \ldots, 2^k-2$ oraz $i \neq i_1$ wybieramy oznaczamy symbolem χ_{i_2}
wartość największą. Podzbiór, któremu odpowiada ta największa wartość

4) Sprawdzamy, czy zachodzi nierówność

$$\bigwedge_j M^2(\chi_{i_2}) > M_j^2(\chi_{i_2})$$

(χ_i).

Jeżeli tak, to za optymalny zbiór zmiennych objaśniających przyjmujemy zbiór χ_{i_2} Jeśli nie, to znów wybieramy podzbiór χ_{i_3}itd.

Model, który otrzymamy w wyniku opisanej metody jest modelem o największej wartości miernika N spośród modeli koincydentnych.

Poniżej przedstawimy przykład ilustrujący drugą z zaproponowanych metod.

Przykład

Niech $\chi = \{X_1, X_2, X_3\}$. Załóżmy, że dla zmiennych objaśniających i zmiennej objaśnianej macierze R i R_0 mają postać:

$$R = \begin{vmatrix} 1 & 0,3 & 0,5 \\ 0,3 & 1 & 0,8 \\ 0,5 & 0,8 & 1 \end{vmatrix} \qquad R_0 = \begin{vmatrix} 0,6 \\ 0,2 \\ 0,3 \end{vmatrix}$$

Zbiór χ ma 7 następujących podzbiorów:

$\chi_1 = \{X_1, X_2, X_3\}$, $\chi_2 = \{X_1, X_2\}$, $\chi_3 = \{X_1, X_3\}$, $\chi_4 = \{X_2, X_3\}$, $\chi_5 = \{_1\}$,
$\chi_6 = \{X_2\}$, $\chi_7 = \{X_3\}$.

Dla każdego podzbioru zbioru χ obliczamy wartość miernika N i otrzymujemy:

N $(\chi_1)=0,08$, N $(\chi_2)=0,30$, N $(\chi_3)=0,25$, N $(\chi_4)=0,04$,
N $(\chi_5)=0,36$, N $(\chi_6)=0,04$, N $(\chi_7)=0,09$.

Spośród wartości N (χ_i) $(i=1, 2, \ldots, 7)$ wybieramy wartość największą czyli

$$N(\chi_{i1}) = N(\chi_5) \gg N(\chi_i) \quad (i=1, 2, \ldots, 7).$$

Ponieważ podzbiór χ_5 ma tylko jedną zmienną objaśniającą X_1, więc model opisywany przez tą zmienną jest koincydentny (model o jednej zmiennej objaśniającej zawsze jest koincydentny por. ([2]). Stąd za otpymalny zbiór zmiennych objaśniających przyjmujemy zbiór χ_5.

LITERATURA

[1] A. B o r o w i e c k i, *Metody doboru zmiennych a problem koincydencji*, Praca doktorska, SGPiS 1983.

[2] A. B o r o w i e c k i, J. K a l i s z y k, M. K o l u p a, *Koincydencja i efekt katalizy w liniowych modelach ekonometrycznych*, PWN, Warszawa 1986.

[3] A. B o r o w i e c k i, M. P ę c z k o w s k i, *Ocena jakości modelu ekonometrycznego*, PWE, Warszawa 1984.

[4] T. G r a b i ń s k i, S. W y d y m u s, A. Z e l i a ś, *Metody doboru zmiennych w modelach ekonometrycznych*, PWN, Warszawa 1982.

[5] M. G r u s z c z y ń s k i, M. K o l u p a, E. L e n i e w s k a, G. N a p i ó r k o w s k i, *Miary zgodności, metody doboru zmiennych, problemy współliniowości*, PWN, Warszawa 1979.

[6] B. G u z i k, *Propozycja kryterium zmodyfikowanego współczynnika determinacji dla doboru zmiennych objaśniających do modelu ekonometrycznego*, „Przegląd Statystyczny" 1979, z. 1/2.

[7] Z. H e l l w i g, *Efekt katalizy w modelu ekonometrycznym, jego wykrywanie i usuwanie*, „Przegląd Statystyczny" 1977, z. 2.

[8] M. K o l u p a, *Macierze brzegowe w badaniach ekonometrycznych* PWE, Warszawa 1982.

[9] M. K o l u p a, *O kryterium dotyczącym badania koincydencji danej zmiennej objaśniającej*, „Przegląd Statystyczny" 1986, z. 4.

[10] J. R o g o w s k i, *Warunek konieczny i dostateczny na zachodzie koincydencji w modelu ekonometrycznym oraz o pewnym sposobie doboru zmiennych objaśniających*, „Przegląd Statystyczny" 1980, z. 3/4.

[11] H. S m i t h, *Analiza regresji stosowana*, PWN, Warszawa 1973.

SUMMARY

The paper presents a proposition of such two methods of the choice of variables to the linear one-equation econometric model which ensure the model coincidence, a minimum intensification of the catalysis effect, a strong correlation between the explanatory variables and the explained variable as well as a poor correlation between the explanatory variables.